Received: 17 May 2023 | Revised: 20 November 2023 | Accepted: 28 November 2023

DOI: 10.1002/jsp2.1309

RESEARCH ARTICLE

JOR Spine OPEN ACCESS

SSR1 and CKAP4 as potential biomarkers for intervertebral disc degeneration based on integrated bioinformatics analysis

Danqing Guo[1,2] | Min Zeng[3] | Miao Yu[4] | Jingjing Shang[4] | Jinxing Lin[4] | Lichu Liu[1] | Kuangyang Yang[1] | Zhenglin Cao[4]

[1]Institute of Orthopaedics and Traumatology, The 8th Clinical Medical College of Guangzhou University of Chinese Medicine, Foshan, Guangdong, China

[2]Guangzhou University of Chinese Medicine the First Affiliated Hospital, Guangzhou, 中国

[3]Pathology Department, The 8th Clinical Medical College of Guangzhou University of Chinese Medicine, Foshan, Guangdong, China

[4]Spinal Surgery Department, The 8th Clinical Medical College of Guangzhou University of Chinese Medicine, Foshan, Guangdong, China

Correspondence
Danqing Guo, Institute of Orthopaedics and Traumatology, The 8th Clinical Medical College of Guangzhou University of Chinese Medicine, No. 6 Qinren Road, Chancheng District, Foshan, 528000, Guangdong, China.
Email: dr.danielaguo@gmail.com

Zhenglin Cao, Spinal Surgery Department, The 8th Clinical Medical College of Guangzhou University of Chinese Medicine, No. 6 of Qinren Road, Foshan, 528000, Guangdong, China.
Email: caozl@fshtcm.com.cn

Funding information
The Grants of Peak Climbing Project of Foshan Hospital of Traditional Chinese Medicine (CN), Grant/Award Numbers: 2022000189, 202200087, 202000190, 202000206

Abstract

Background: Intervertebral disc degeneration (IDD) is a significant cause of low back pain and poses a significant public health concern. Genetic factors play a crucial role in IDD, highlighting the need for a better understanding of the underlying mechanisms.

Aim: The aim of this study was to identify potential IDD-related biomarkers using a comprehensive bioinformatics approach and validate them in vitro.

Materials and Methods: In this study, we employed several analytical approaches to identify the key genes involved in IDD. We utilized weighted gene coexpression network analysis (WGCNA), MCODE, LASSO algorithms, and ROC curves to identify the key genes. Additionally, immune infiltrating analysis and a single-cell sequencing dataset were utilized to further explore the characteristics of the key genes. Finally, we conducted in vitro experiments on human disc tissues to validate the significance of these key genes in IDD.

Results: we obtained gene expression profiles from the GEO database (GSE23130 and GSE15227) and identified 1015 DEGs associated with IDD. Using WGCNA, we identified the blue module as significantly related to IDD. Among the DEGs, we identified 47 hub genes that overlapped with the genes in the blue module, based on criteria of $|logFC| \geq 2.0$ and $p.adj < 0.05$. Further analysis using both MCODE and LASSO algorithms enabled us to identify five key genes, of which CKAP4 and SSR1 were validated by GSE70362, demonstrating significant diagnostic value for IDD. Additionally, immune infiltrating analysis revealed that monocytes were significantly correlated with the two key genes. We also analyzed a single-cell sequencing dataset, GSE199866, which showed that both CKAP4 and SSR1 were highly expressed in fibrocartilage chondrocytes. Finally, we validated our findings in vitro by performing real time polymerase chain reaction (RT-PCR) and immunohistochemistry (IHC) on 30 human disc samples. Our results showed that CKAP4 and SSR1 were upregulated in degenerated disc samples. Taken together, our findings suggest that CKAP4 and SSR1 have the potential to serve as disease biomarkers for IDD.

KEYWORDS
immune infiltration, intervertebral disc degeneration, single-cell sequencing, WGCNA

1 | INTRODUCTION

Low back pain has become a significant healthcare concern due to its impact on daily life and medical insurance costs.[1–3] Intervertebral disc degeneration (IDD) is the most common cause of low back pain,[4] and current management options for IDD, including conservative treatment and surgery, cannot cure disc degeneration. Genetic factors have been identified as playing a role in IDD, and recent research has focused on identifying potential biomarkers for the condition.[5–8]

Weighted gene coexpression network analysis(WGCNA) is a powerful tool for identifying genes driving major signaling pathways.[9] Previous studies by Li et al.,[6] Wang et al.,[7] and Ma et al.[10] identified potential biomarkers for IDD using WGCNA method, but their findings may have limitations. Li et al. identified SIRT7, NTRK2, and CHI3L1 from two datasets with different sample origins, which may affect the accuracy and reliability of their results. Ma et al. reported ASAP1-IT1 and SERINC2 as critical genes related to IDD from a dataset of blood samples, which may not accurately reflect gene expression in disc tissues. Wang et al. identified SMIM1 and SEZ6L2 from a dataset with annulus cells exposed to TNF-α, which only targeted inflammation-related genes. To address these limitations and explore degeneration-related genes, we included datasets from disc tissues graded using the same classification system. This approach allowed us to identify key genes related to IDD with greater accuracy and reliability. In addition to WGCNA, the emergence of single-cell RNA sequencing (scRNA-seq) has further enhanced our understanding of cell heterogeneity and the mechanisms of IDD development. By identifying unique chondrocyte subsets during the process of IDD and signature transcription factors of annulus fibrosis (AF) cells and nucleus pulposus (NP) cells[11,12] scRNA-seq has provided valuable insights into gene function and distribution in cell subtypes.

Moreover, the immune response has been reported to play a critical role in IDD development,[13–15] largely depending on the mediation of inflammation. While the intervertebral disc is an immune-privileged organ with NP isolated from the immune system by the blood-tissue barrier, damage to this barrier can lead to exposure of NP to the immune system.[13,16] This exposure can trigger the release of cytokines and chemokines that recruit immune cell migration to mediate inflammation, driving catabolism and degrading extracellular matrix.[17] Thus, understanding the regulation of immune response is necessary for IDD treatment.

In this present study, we aimed to identify IDD-related key genes via multiple bioinformatics methods, including WGCNA, and further explore the mechanisms of the key genes regulating immune response and identifying which kind of cell subtype they mainly expressed in through scRNA-seq. Furthermore, we validated the key genes in vitro with human disc tissues. Our work provides a novel and trustworthy perspective for potential biomarkers for IDD, which could ultimately lead to improved diagnostic and therapeutic strategies for this common condition.

2 | MATERIALS AND METHODS

2.1 | Data collection

From the National Center of Biotechnology Information (NCBI) Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) database, the datasets of GSE23130 and GSE15227 including a total of 38 disc tissue samples were downloaded, and based on GPL1352 platform. Samples from GSE23130[18] and GSE15227[19] were obtained via the National Cancer Institute Cooperative Tissue Network (CHTN) as well as surgical disc procedures performed on patients with herniated discs and degenerative disc disease. The degenerative degree of all samples in both datasets was graded according to the Thompson grading criteria.[20] The method used to extract the total RNA of GSE23130 involved homogenizing disc tissues in TRIzol reagent. On the other hand, the RNA of GSE15227 was obtained from cultured cells that were isolated from disc tissues. In this study, we included gene expression data from IDD samples graded as Thompson grades IV and V and control samples graded as I and II for data mining. Twenty-two samples (11 control samples and 11 IDD samples) were selected for differential expression genes (DEG) analysis. Meanwhile, the gene expressions for the 22 samples were collected for WGCNA analysis. Besides, GSE70362 based on the GPL17810 platform including 24 nucleus pulposus cell samples and 24 annulus fibrosus cell samples, and scRNA-seq dataset GSE199866 including 4 samples were used as validation datasets.

2.2 | Clinical samples

Disc tissues for in vitro experiments were collected from patients diagnosed with spinal fractures or lumbar degenerative diseases who needed discectomy undergoing spinal surgeries. The degeneration of discs was graded according to the Pfirrmann classification.[21] Samples from patients with spinal fractures and dislocation were regarded as normal control graded as Pfirrmann I or II. All specimens were frozen in liquid nitrogen instantly after they were removed from the body and stored at −80°C till they were used for further experiments.

2.3 | Differential gene expression and principal component analysis

The R software (version 3.6.3) was used with the Limma package version 3.42.2.[22] It was considered differentially expressed when genes with fold change (Log_2FC) >1 (upregulated) or <−1 (downregulated) had adj. p values of 0.05 or greater. Using hierarchical clustering and R software, heatmaps and volcano plots were constructed to visualize differentially expressed gene expression profiles. A principal

component analysis (PCA) was performed on the samples based on their gene expression data.

2.4 | Coexpression network construction

WGCNA R package (version 1.70-3) was used to construct the coexpression network for the expression data.[9] The steps were as follows: first, we used the hclust function to analyze the sample cluster for removing outliers; Second, a pickSoftThreshold function was used to create a scale-free network[23,24]; Third, create a topological overlap matrix (TOM) transformed from adjacency matrix which was converted from similarity matrix by calculating the weighting coefficient, β; Fourth, we set the minimal module size of 30 genes for the dynamic tree cutting algorithm and then every each gene was assigned into individual coexpression modules of which correlation above 0.25 was merged.[25]

2.5 | Identification of significant modules

The modules showing the closest correlation coefficient with IDD were determined as IDD-related modules by calculating the Pearson correlation coefficient among the module eigengenes (MEs) of all modules and IDD. Additionally, the gene significance (GS) was calculated between the gene expression levels and IDD. All genes' GS values were calculated based on their mean absolute values.[9] A higher mean absolute value indicates a stronger relationship between a module and IDD.

2.6 | Identification of hub genes

The hub genes play a crucial role in the functioning of the module. Module memberships (MMs) are used to determine correlations between genes and modules as part of hub gene screening.[9] Gene expression profiles and MEs for each module are correlated to determine the MM. Connectivity within modules is strongly associated with the MM measure.[9] Intramodular hub genes with high MM values tend to be highly connected. The higher the MM value of the gene, the more significant the correlation between the gene and a module. Hub genes were identified by detecting genes associated with high within-module connectivity (|MM| >0.8) and strong Pearson's correlation (|GS| >0.2) with specific clinical traits.[26] Therefore, in this study, we used (|MM| >0.8 and |GS| >0.7) as a stricter cut-off. To further narrow down the range of hub genes, we extracted the common targets of hub genes in the module and the DEGs were extracted using Venny 2.1 (https://bioinfogp.cnb.csic.es/tools/venny/).

2.7 | Gene ontology, KEGG pathway enrichment, and MCODE analysis

A clusterProfiler R package (version: 3.14.3) was also used for enrichment analysis with Log_2FC value of each gene.[27] This R package used the default parameters for clusterProfiler. Thresholds of p.adj < 0.1 and q-value < 0.2 were chosen, respectively. Org. Hs.eg.db package was used to convert Uniprot IDs, and the z-scores were calculated by GOplot package (version 1.0.2).[28]

2.8 | Identification of key genes with diagnostic value

Two methods were used to identify key genes. On the one hand, a protein–protein interactions (PPIs) network was generated by Metascape (metascape.org), and the MCODE algorithm was then applied to identify densely connected proteins as key genes. The visualization of the PPI network was edited by Cytoscape (version 3.9.0). On the other hand, the hub genes were analyzed using the lease absolute shrinkage and selection operator (LASSO) logistic regression[29] with Glmnet package in R. All 38 samples from datasets of GSE23130 and GSE15227 were input, with the control group consisting of samples graded as Thompson I, II, and III, and the IDD group consisting of samples graded as IV and V. The parameter settings were family = "binomial,"[30] cross-validation = 8 fold with "type. measure = auc," lambda = 10, and a training set size of 0.7. Additionally, the diagnostic value of the key genes in distinguishing IDD from control samples was assessed by calculating the area under the receiver operating characteristic (ROC) curves using the pROC package in R.[31] Furthermore, the key genes were validated in GSE70362.

2.9 | Correlation analysis with infiltrating immune cells

An analysis of single sample gene sets enrichment (ssGSEA) was performed to investigate immune infiltration between control and IDD samples. A total of 22 samples from GSE23130 and GSE15227 datasets were included in analysis. A significance level of $p < 0.05$ was set for the sample filter. The "split-violplot" package was applied to visualize the correlation of immune cells between IDD and control samples. The "pheatmap" package was used to plot the heatmap of immune infiltration in the samples. Moreover, the correlation between the immune infiltration cells and IDD-related key genes was evaluated using the "psych" and "ggcorrplot" packages. Additionally, the "ggplot2" package was used for the analysis and visualization of the results, with a significance level of $p < 0.05$.

2.10 | Analysis of key genes in cell clusters based on scRNA-seq

GSE199866 contains four samples of non-degenerating and degenerating NP and annulus fibrosis (AF) cells (NPN, NPD, AFN, AFD), with 14,001 cell. The Seurat (v4.1.2) R package was used for processing the data and downstream analysis. Quality control was performed to exclude the cells with UMI counts smaller than the lower quartile or larger than the upper quartile, and mitochondrial gene percentage

<5%. Then, PCA was performed to reduce the data dimensionally. The first 10 principal components from the scale data were used for clustering and tSNE representations. With a resolution of 0.5, clusters were identified in Seurat using the FindClusters function. The expression of key genes in different kinds of cell clusters was visualized using the ggplot2 package. The annotation of cell subtypes used specific cell markers obtained from the CellMarker website (http://biocc.hrbmu.edu.cn/CellMarker/).

2.11 | RNA isolation and RT-PCR experiment

Total RNA was extracted by TRIZOL method (Trizol, Invitrogen, #15596026). Synthesis of cDNA was used the RevertAid first-strand cDNA synthesis kit (catalog number #K1621). GAPDH was utilized as an internal control. The Human CKAP4 primer sequences were as follows: (forward) 5′-CAGCCGGATCAGCGAAGT-3′; (Reverse) 5′-TGTGAAGATGGCGATGTTGT-3′. The Human SSR1 primer sequences: (forward) 5′-CTGCTTCTCTTACTCGTGTTCC-3′; (Reverse) 5′-TCTTCTTCTACCTCGGCTTCAT-3′. The Human GAPDH primer sequences: (forward) 5′-GTCCACTGGCGTCTTCACC-3′; (Reverse) 5′-CATGAGTCCTTCCACGATACCAA-3′. RT-PCR was performed using the PowerUp™ SYBR™ Green Master Mix kit (catalog number #A25741) and ran on the Applied Biosystems 7500 system. The running conditions followed the manufacturer's instructions. The relative expression level of genes was normalized to GAPDH and calculated by using the $2^{-\Delta CT}$ method.

2.12 | Histologic analysis

The specimens were collected from spinal surgeries, and immediately fixed in 10% formalin for 48 hours. Subsequently, they were dehydrated and embedded in paraffin. The steps for immunohistochemistry (IHC) were as follows: the sections were (1) deparaffinized and rehydrated and then microwaved in 0.01 M sodium citrate or PH = 9 tris-EDTA for 15 minutes, (2) incubated with 3% hydrogen peroxide for 10 min to block endogenous peroxidase activity, (3) followed with 5% bovine serum albumin for 30 minutes to block nonspecific binding sites, (4) followed with primary antibody (CKAP4, SSR1, 1:200, Proteintech) overnight at 4°C, (5) and finally incubated with an HRP-conjugated secondary antibody (GE) and counterstained with hematoxylin. Three sections from each specimen were minimal for performing tests. Images from IHCs were measured for integrated optical density (IOD) using Image J (version 1.52). The average optical density (AOD) was calcultaed using the formula: AOD = IOD/Area.

2.13 | Statistics

R or Graph Pad Prism 7 was used for statistical analysis with data presented in mean ± standard deviation (SD). Plots were generated using the corresponding R package or Graph Pad Prism 7. One-way ANOVA followed by Dunnett's test was used for comparisons among multiple groups, while a t-test was used for comparisons between two groups. A p value of <0.05 was considered statistically significance for identifying differences.

3 | RESULTS

3.1 | Identification of DEGs

One thousand and fifteen DEGs in total were identified in GSE23130 and GSE15227. Among them, 98 genes were downregulated, and 917 were upregulated based on the criterion of p.adj value <0.05 and |log2FC| >1. A volcano plot is shown in Figure 1A, where the upregulated genes are in red, and the downregulated genes are in blue. The heatmap of the DEGs (Figure 1B) displays the expression patterns of genes that are significantly different between normal and IDD groups. Additionally, PCA demonstrated that the 11 normal disc tissue samples were clustered separately from the other 10 IDD samples (Figure 1C). Twenty-two disc samples were clustered into a tree map with corresponding group classification shown in Figure 1D.

3.2 | Identification of IDD-related module-trait hub genes

A coexpression network was constructed using a weighting coefficient of $\beta = 10$, which was selected based on log(k) and log[$P(k)$] (Figure S1A,B). The network exhibited a scale-free property with a correlation coefficient of >0.86 (Figure 1C,D). Hierarchical clustering was performed to assign genes into different clusters by setting the minimum number of genes as 30 and deepSplit as 2, resulting in 156 modules (Figure 2A). These modules were further merged into 66 modules based on average linkage hierarchical clustering with a correlation coefficient >0.75 (Figure 2A). The correlation between modules and degeneration degree was displayed in Figure S2. The blue module showed the highest significant correlation with IDD ($R = 0.69$, $p = 3e-04$) and the largest mean absolute gene significance value (cor = 0.72, $p = 1.5e-200$, Figure S3), making it the highly related module (Figure 2B). Next, under |MM| >0.8 and |GS| >0.7, 56 genes were collected from the blue module. Besides, 508 genes were obtained when $|Log_2FC| \geq 2.0$ and p.adj < 0.05 were employed in DEGs. As a result, 47 hub genes were further screened out from the two groups.

3.3 | Hub gene functional enrichment analysis

The analysis was based on the z-score and Log_2FC of each hub gene and performed with clusterProfiler R package (version: 3.14.3), org.Hs.eg.db package (version: 3.10.0), and GO plot package (version: 1.0.2).

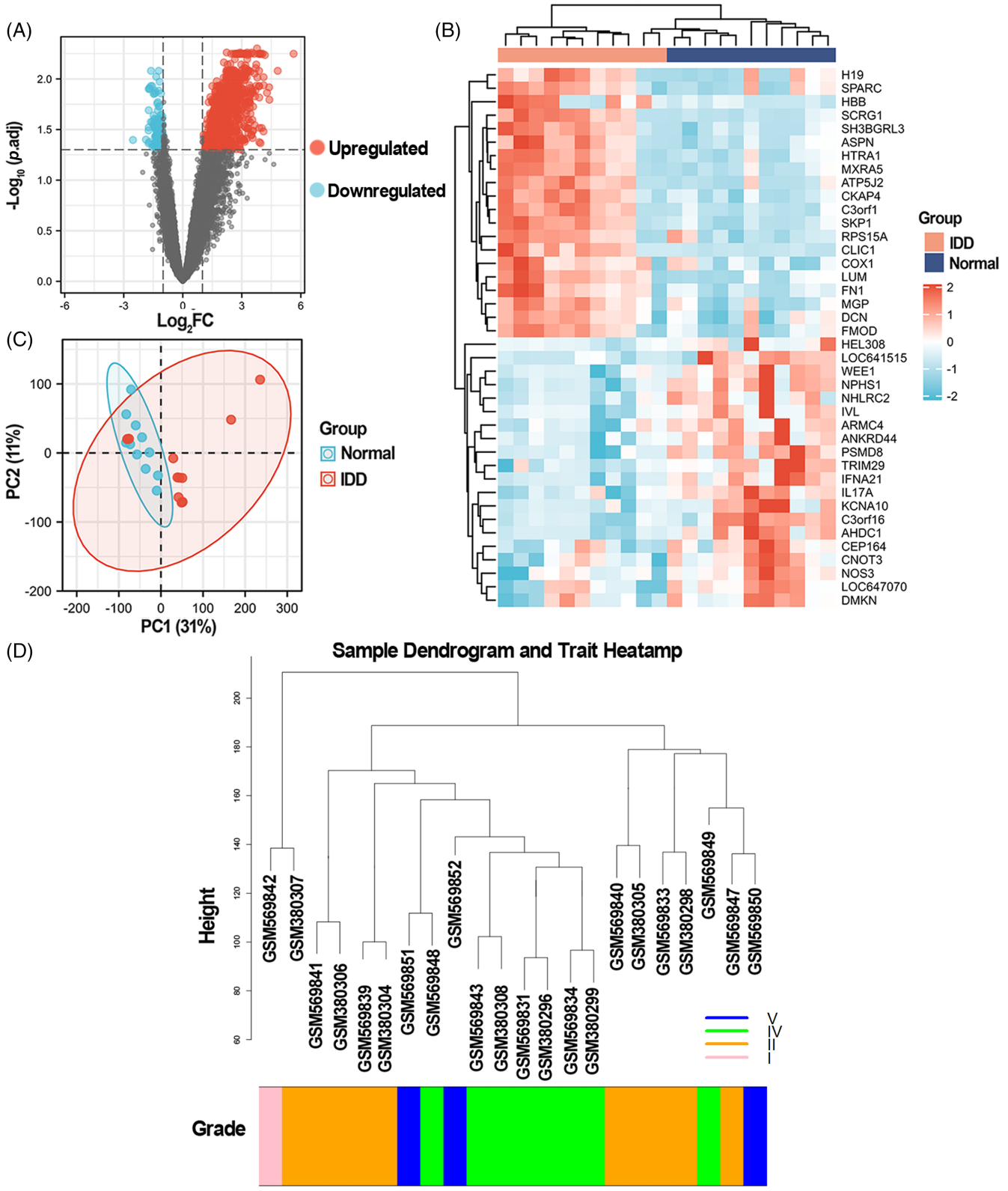

FIGURE 1 Differential mapping analysis of DEGs. (A) The volcano plot of 1015 differentially expressed mRNAs between IDD and normal samples (|Log_2FC| >1, *p*.adj. <0.05, 917 upregulated genes in red and 98 downregulated genes in blue dots). (B) The heatmap of the top 20 out of 1015 DEGs between IDD and Normal tissues (|Log_2FC| >1, *p*.adj. <0.05). Red indicates high expression and blue indicates low expression. (C) PCA plot of 1015 DEGs. (D) Twenty-two samples clustered into a tree with corresponding grade classification in the GSE23130 and GSE15227 datasets (no identified outliers).

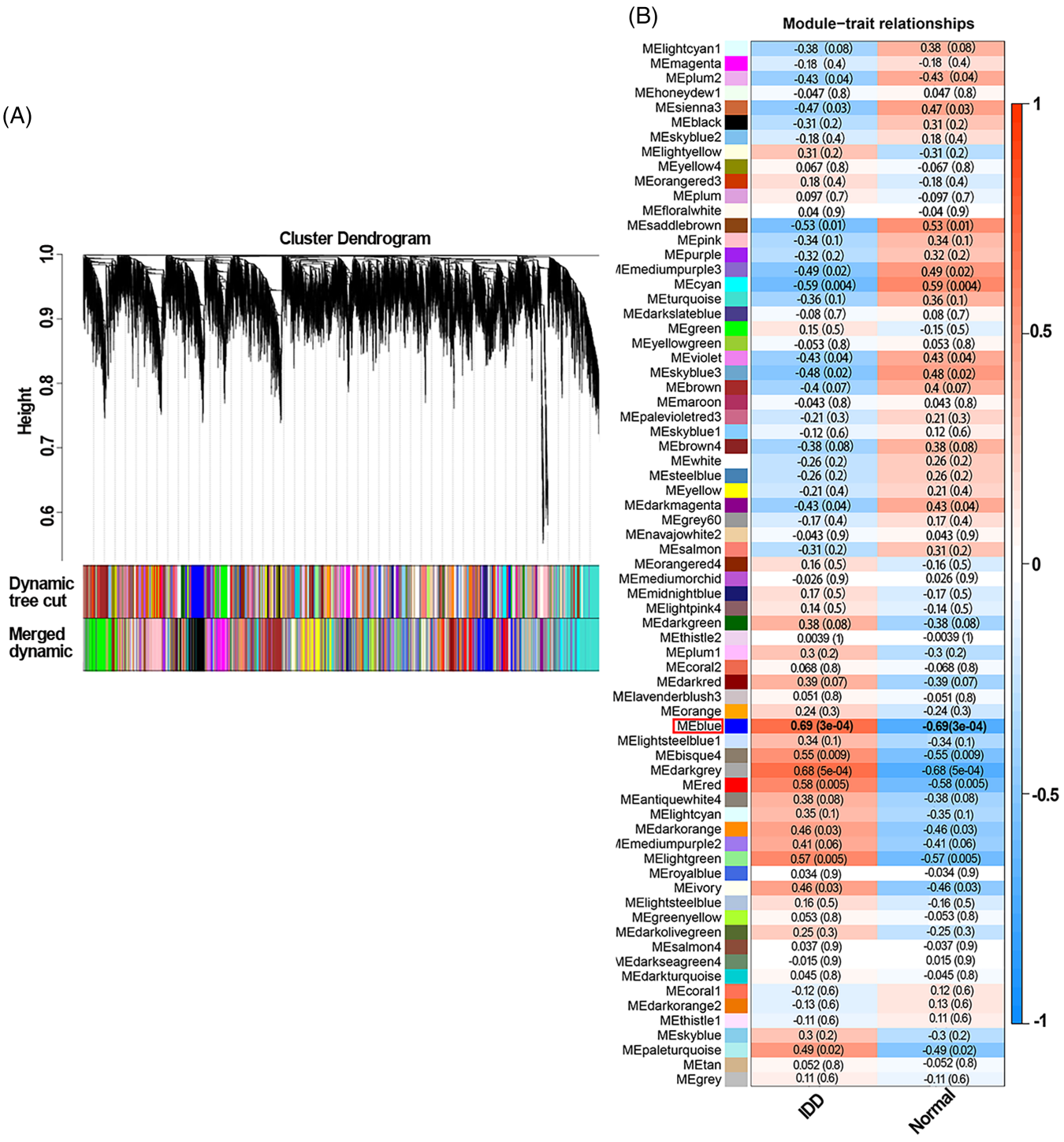

FIGURE 2 Network heatmap of disc samples. (A) Clustering dendrograms of all DEGs. The clustering dendrograms of all DEGs were created using dissimilarity based on topological overlap. Additionally, the dendrograms were annotated with assigned module colors. In total, 66 coexpression modules were formed and represented using different colors. (B) Module-trait relationship heatmap. The association between modules and traits. A trait is associated with each column, and an ME is associated with each row. The correlation coefficient value is in rectangles, and the corresponding *p*-value is in brackets. The blue module (enclosed with a red box) shows the highest significant correlation with IDD.

A total of 34 cellular components, 5 biological processes, 63 molecular functions, and 1 KEGG pathway were enriched and presented in Table S1. The top 16 enriched genes (SSR4, SSR1, SKP1, TCFL2, FOXP1, GNB1, CKAP4, CD63, CD164, SEC13, PPP2CB, DYNC1LI2, GTF2A2, DAXX, TAF7, SET) were displayed in the Figure 3A. The most enriched functions included RNA polymerase II-specific DNA-binding transcription factor binding, DNA-binding transcription factor binding, nuclear receptor binding, and histone binding. The only pathway involved was protein processing in the endoplasmic reticulum. The results were visualized in Figure 3B.

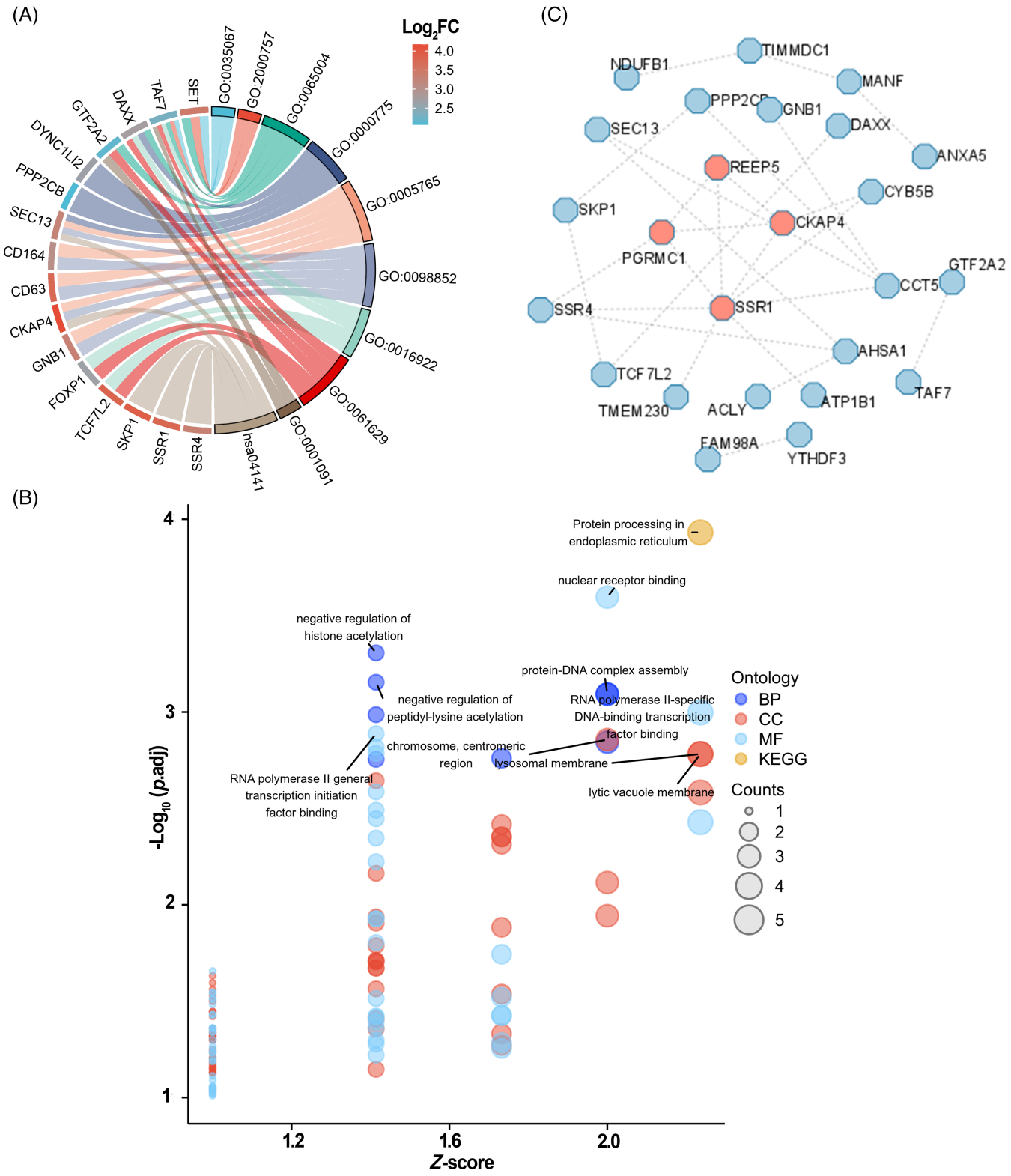

FIGURE 3 Functional enrichment analysis and key genes. (A) GO Chord plot shows the representative genes, GO terms, and one pathway corresponding to hub genes. (B) The bubble map shows the biological process, cellular component, molecular function, and KEGG pathway of 47 hub genes. (C) MCODE algorithm was applied to a PPI network, identifying four key genes.

The PPI network analysis identified four proteins, CKAP4, SSR1, PGRMC1, and REEP5, as densely connected using the MCODE algorithm (Figure 3C). Additionally, SSR1 and BNC2 were identified as key genes using the LASSO algorithm (Figure 4A,B).

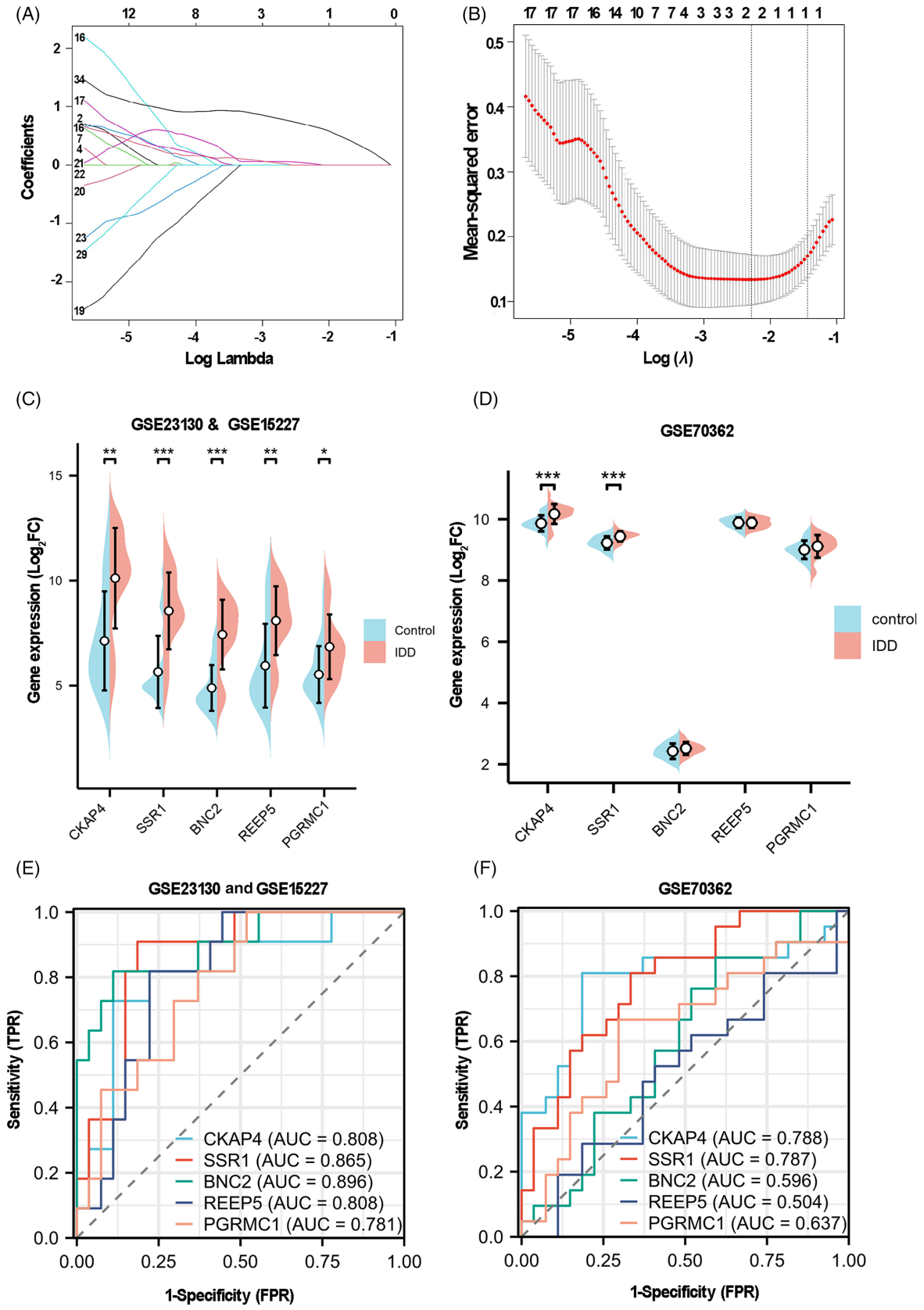

FIGURE 4 Identification of key genes with great diagnostic value for IDD. (A) Lambda value of the 47 hub genes in the LASSO model. (B) The most proper Lambda value in the LASSO model. (C) The expression levels of the five key genes in GSE23130 and GSE15227 datasets. (D) Validation of the five key gene expressions in GSE70362 dataset. (E) ROC curves for the five key genes in GSE23130 and GSE15227 datasets. (F) ROC curves for the five key genes in GSE70362 dataset for validation.

3.4 | Identification and diagnostic value of key genes

To validate the predictive power of these five key genes in IDD samples, we used an independent dataset, GSE70362. In comparison to the significant differential expressions observed between control and IDD samples in the GSE23130 and GSE15227 datasets (Figure 4C), and AUC values above 0.7 (Figure 4E), only CKAP4 and SSR1 exhibited differential expressions (Figure 4D) and had AUC values above 0.7 in the GSE70362 dataset (Figure 4F).

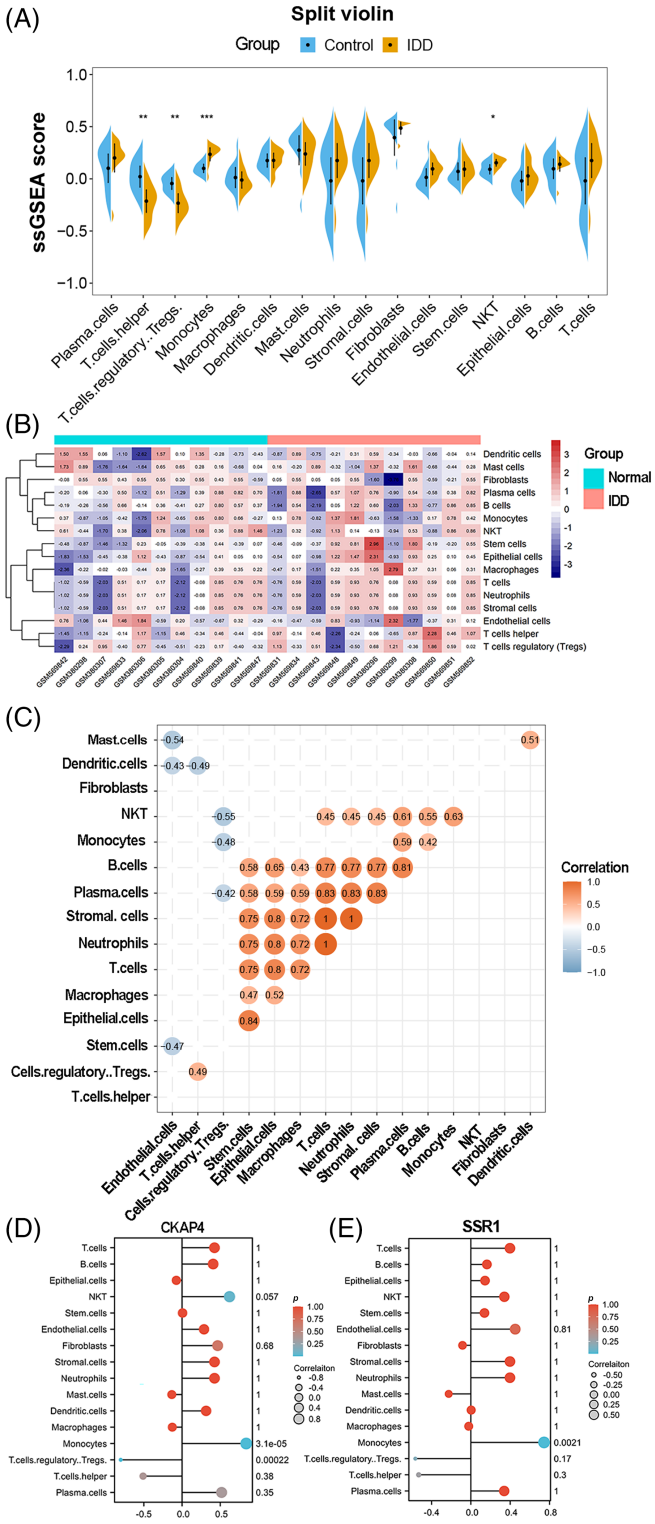

FIGURE 5 Correlation with infiltrating immune cells. (A) Split-violin plot shows the differential correlation of immune cells between the control and IDD samples. (B) Heatmap for the differential correlation of immune cells between the control and IDD samples. (C) The dot plot shows the correlation among 22 kinds of infiltrating immune cells. (D) The lollipop plot shows the correlation between the CKAP4, SSR1, and immune cells, respectively. $^{*}p < 0.05$ and $^{**}p < 0.01$.

3.5 | Correlation with immune infiltration

To explore the correlation with infiltrating immune cells, ssGSEA scoring was performed to compare the composition and proportion of infiltrating immune cells between control and IDD samples in the GSE23130 and GSE15227 datasets. As shown in Figure 5A, T cells helper and T cells regulatory were significantly more correlated with control samples ($p < 0.01$ and $p < 0.05$, respectively), while monocytes showed a greater correlation with IDD samples ($p < 0.05$). The correlation value of each sample against every kind of immune cell was shown in Figure 5B.

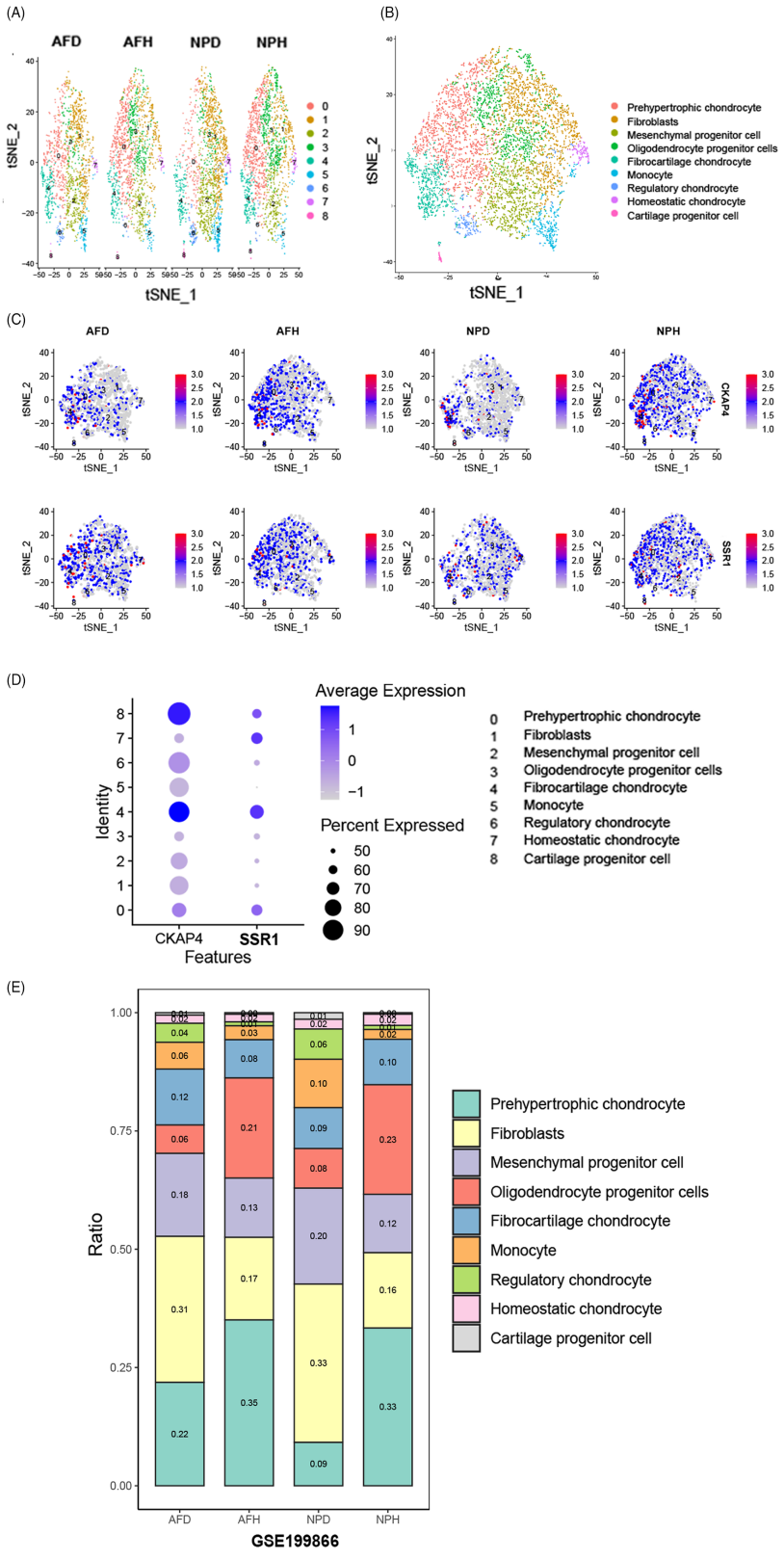

FIGURE 6 Identification of the cell subtypes that two key genes are mainly expressed in. (A) Unsupervised tSNE clustering shows the change in cell distribution of the nine clusters for AFD, AFH, NPD, and NPH. (B) Cell annotations for the nine clusters. (C) Expression distributions of the two key genes in different cell clusters attributed to different samples. (D) The average expression of the two key genes in different cell subtypes. (E) The proportions of the nine cell subtypes attributed to different samples.

Moreover, by analyzing the correlation of each kind of infiltrating immune cells, we found a significant positive correlation between T cells helper and T cell regulatory and a significant negative correlation between monocytes and T cell regulatory (Figure 5C). Further analysis revealed a significant positive correlation between CKAP4 and SSR1 expressions with monocytes (Figure 5D,E).

3.6 | The characteristics of the key genes in cell clusters

To determine the cell type expressing these key genes, we introduced a single-cell RNA-seq dataset, GSE199866. All cells were clustered into nine clusters (Figure 6A), and each subtype was annotated (Figure 6B) as follows: prehypertrophic chondrocyte, fibroblasts, mesenchymal progenitor cell, oligodendrocyte progenitor cells, fibrocartilage chondrocyte, monocyte, regulatory chondrocyte, homeostatic chondrocyte, and cartilage progenitor cell.

As shown in Figure 6C,D, CKAP4 was significantly and differentially expressed highest in fibrocartilage chondrocytes and second in cartilage progenitor cells. Moreover, the proportion of fibrocartilage chondrocyte was highest (0.12) in the AFD sample (Figure 6E) compared to the other three samples; and the monocyte proportion was highest in NPD (0.10) and second in AFD (0.06) sample. SSR1 was not included in the top 2000 hypervariable genes which indicated the expression was not so different among cells as the other 2000 genes, however, it still showed the highest expression in fibrocartilage chondrocytes.

3.7 | Key gene verification

3.7.1 | Clinical samples

From November 2022 to October 2023, a total of 30 human disc tissues were obtained, with an equal number of male and female donors (15 each). The donors had an average age of 56 ± 16.5 years, ranging from 31 to 83 years old. The patient information was listed on Table S2.

3.7.2 | Validation of mRNA and protein expressions

Compared to disc tissues graded as Pfirrmann II, those graded as Grade III and IV showed a significant increase in the mRNA expression of CKAP4 ($p = 0.0496$ and $p = 0.0085$, respectively), while Grade V showed a trend increase without statistical significance ($p = 0.1597$). Additionally, the protein expression of CKAP4 increased significantly in Grade III ($p = 0.0328$), IV ($p = 0.0270$), and V ($p = 0.0397$) as determined by immunohistochemistry. Furthermore, the mRNA expression of SSR1 was significantly increased in Grade IV ($p = 0.0061$) and Grade V ($p = 0.0019$), but not in Grade III ($p = 0.4778$). The protein expression

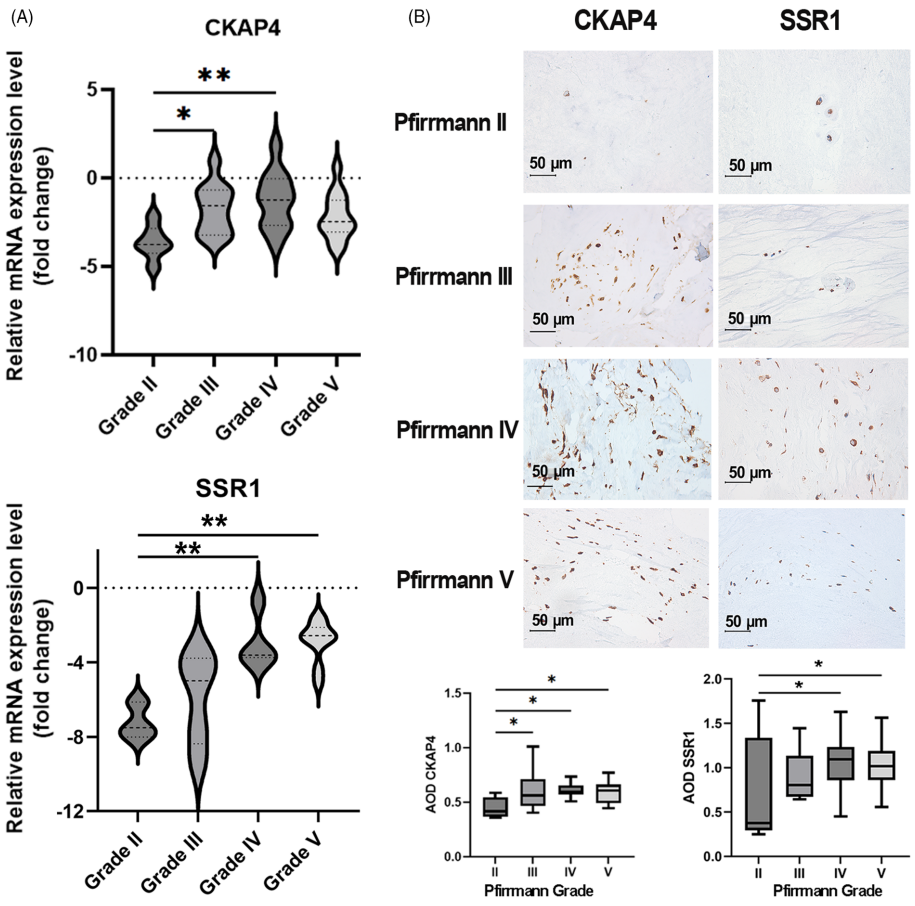

FIGURE 7 The mRNA and protein expression of CKAP4 and SSR1. (A) The mRNA expression of CKAP4 and SSR1 was analyzed using RT-PCR. The relative mRNA expression was normalized against GAPDH and presented as fold change. A significance level of $p < 0.05$ was used to determine statistical significance. For CKAP4, there were three samples in the Pfirrmann II group, eight in the Pfirrmann III group, seven in the Pfirrmann IV group, and nine in the Pfirrmann V group. For SSR1, there were three samples in the Pfirrmann II group, and six samples per group from Pfirrmann III to V. (B) IHC staining of NP tissues. Representative CKAP4 and SSR1 IHC staining of disc tissues graded from Pfirmann II to V. (magnification: 40×, scale bar =50 μm). For CKAP4, $n = 3$ in group Pfrrimann II, $n = 5$ in group Pfrrimann III, $n = 4$ in both group Pfrrimann IV and V; For SSR1, $n = 3$ per group.

of SSR1 was significantly increased in Grade IV ($p = 0.0265$) and V ($p = 0.0227$), but not in Grade III ($p = 0.4209$). The results have been presented in Figure 7A,B.

DISCUSSION

IDD is a pathological process that occurs with aging, and genetic factors contribute to 70% of its occurrence.[32,33] To better understand the underlying mechanisms, we applied bioinformatics methods to identify biomarkers associated with IDD. We analyzed 1015 DEGs from 22 samples using WGCNA to calculate correlations between gene expression profiles and IDD. We excluded disc samples graded as Thompson III due to their uncertainty in morphological change,[20] resulting in a total of five modules associated with IDD. The blue module was found to be the most strongly correlated with IDD, and we identified 47 hub genes for further analysis. Notably, only the pathway of "protein processing in endoplasmic reticulum" was enriched in the 47 genes.

The endoplasmic reticulum (ER) is quite sensitive to extracellular stimuli in disc cells because matrix proteins are abundantly synthesized and secreted. Liao et al.[34] have reported that human NP samples with varying degrees of degeneration showed a positive correlation between the upregulation of ER stress-related apoptosis genes and Pfirrmann grades of IDD, implicating that ER stress contributes to IDD. Last year, Luo et al.[35] prominently reported that *O*-linked β-*N*-acetylglucosaminylation transferase regulated IDD by targeting FAM134B-mediated ER-phagy, which was published in *Nature*. Therefore, the protein processing pathway in ER is worthy of further investigation for potential therapeutic targets.

In addition, endoplasmic reticulum stress has been disclosed in regulating immune responses,[36,37] and the relationship between IDD and the immune response is very close.[13–15] Therefore, we further investigated the correlation of the 22 samples in the GSE23130 and GSE15227 datasets with immune cell infiltration by the ssGSEA method. As a result, monocytes were found largely infiltrated in IDD samples. Consistently, in the following scRNA-seq dataset analysis, we found monocyte proportion was largest in degenerated NP sample, and second in degenerated AF sample. Similarly, a recent study also found monocytes were aberrantly activated in the late stage of IDD via scRNA-seq.[38] As is known, inflammation plays a key role in the process of IDD. Monocytes, one of the immune cells, as a proinflammatory factor, largely recruited to the site with infection or sterile injury is critical in mediating inflammation resolution and wound repair.[39] Superfluous cytokines and chemokines upregulating in the IDD process produced by native disc cells have been disclosed to mediate catabolic events.[16,40] IL-1 was reported to mediate the catabolic events during IDD,[40] while other cytokines, such as MCP-1, TNF-alfa, and IL-8, diffuse to the surrounding environment triggering inflammation and accelerating immune cells infiltration like T cells (CD4+, CD8+), B cells, macrophages, secondary to AF or chondral endplate rupture.[13,16] These immune cells mediate inflammation and further increase the expression of inflammatory factors,[40] which forms a viscous circle driving catabolism and causing extracellular matrix breakdown.[17] Therefore, monocyte might be an immune therapeutic target for IDD.

Moreover, we further identified the 5 key genes out of 47 hub genes by MCODE and LASSO algorithm. The diagnostic value of these five key genes was tested by another dataset, showing that only SSR1 and CKAP4 had great value to distinguish IDD samples from control samples, which might serve as diagnostic biomarkers. Interestingly, we found these two genes were also highly correlated with monocytes, indicating that they might promote IDD via regulating monocytes.

Signal sequence receptor (SSR) is a translocation-related protein that plays a critical role in translocating secreted proteins. The SSR complex was induced by XBP1/IRE1α pathway and involved in endoplasmic reticulum-associated degradation (ERAD), preferentially bounding to misfolded proteins and discriminating ERAD substrates from correctly folded ones to accelerate the degradation of misfolded proteins.[41] SSR1, one of the four subunits of the SSR complex,[42] has been revealed to participate in several kinds of tumors[43] and is a potential biomarker of Parkinson's Disease.[44] Similarly involved in ER stress pathway, CKAP4 (cytoskeleton-associated protein 4), as well as known as p63, was first discovered as a resident protein of the stable ER-Golgi intermediate compartment,[45] works as a receptor for some ligands, including anti-proliferating factor[46] and tissue plasminogen activator.[47] CKAP4 could regulate the morphology and functions of ER, including the separation of ER sheets adjacent to the nucleus,[48] and anchoring Dicer to the ER to regulate the microRNA pathway and mRNA translation.[49] Besides of main findings of CKAP4 functioning in tumor cells,[50,51] scRNA-seq has been identified as a new marker for activated fibroblasts in diseased hearts of murine and human origin,[52] and silencing of it reduced the VSMCs and aortic calcification by modulating YAP phosphorylation and MMP2 expression in Chronic kidney disease.[53] Recently, CKAP4 has also been revealed to play a role in unfolded protein response (UPR), via regulating the activity of the ER chaperone BIP (binding immunoglobulin protein), which is a key player in the UPR,[54] or interacting with some proteins involved in the UPR, like PERK.[55,56]

In addition, to identify the cell type of the two genes mainly expressed in, a scRNA-seq dataset was used. As a result, CKAP4 was found mainly expressed in fibrocartilage chondrocyte and cartilage progenitor cells, and SSR1 was mainly in fibrocartilage chondrocyte. The fibrocartilage chondrocyte phenotype has been reported mainly attributed to osteoarthritis (OA) development[57,58] by damaging the integrity of the cartilage ECM[59] and enhancing fibrosis.[60,61] And the cartilage progenitor cells (CPCs) were characterized based on the expression of stem-cell-related surface markers, the capability of self-renew, and differentiation to multiple lineages. The migration of CPCs has been disclosed for response to injury, like ECM degradation[62] in healthy cartilage,[63] and regulated by multiple tissue-injury-related signals.[63,64] In degenerated

cartilage sites, CPCs were found to accumulate in areas of tissue repair due to the broken cartilage tidemark and evident neovascularization beneath the cartilage tissue.[65] Therefore, we speculated that CKAP4 and SSR1 might also regulate the differentiation of fibrocartilage chondrocytes and migration of CPCs to affect IDD.

The limitations of this study include the following: first, research in vitro and in vitro is still needed to understand further how these hub genes influence IDD; Second, to strengthen our findings and better clarify the potential mechanisms responsible for these hub genes influencing IDD, more clinical samples will be needed.

4 | CONCLUSIONS

In the present study, the WGCNA method was employed and identified one gene coexpression module with the highest gene significance associated with IDD. Then 47 hub genes were overlapped and collected from both the blue module and DEGs. Subsequently, five key genes were identified by MCODE and LASSO algorithms, however, only CKAP4 and SSR1 were validated with great diagnostic value by a test dataset. In addition, the immune infiltrating analysis showed the monocytes were significantly upregulated in IDD samples and correlated with CKAP4 and SSR1. Moreover, both genes were identified as primarily expressed in fibrocartilage chondrocytes based on a scRNA-seq dataset. Accordingly, the possible mechanisms of the two key genes are believed to potentially impact IDD through the regulation of monocyte migration and fibrocartilage chondrocyte differentiation, which warrants further investigation. Additionally, in vitro verification demonstrated increased expression of CKAP4 and SSR1 in degenerated human disc tissues. Therefore, CKAP4 and SSR1 have been identified as promising therapeutic targets for the treatment of IDD.

AUTHOR CONTRIBUTIONS

Danqing Guo conceived of the presented idea carried out the bioinformatics analysis, drafted the main part of the manuscript and revised it; Min Zeng performed the in vitro experiments and drafted the corresponding method part; Miao Yu verified the analytical methods, carried out the data analysis and patient consent acquisition; Jingjing Shang and Jinxing Lin assisted Min Zeng to carry out the in vitro experiments and Miao Yu to acquire the patient consents; Lichu Liu and Kuangyang Yang provided the findings; Zhenglin Cao supervised the project and substantively revised the manuscript.

ACKNOWLEDGMENTS

I would like to thank Honghong, Liao who provided purely technical help, and Home for Researchers (https://www.home-for-researchers.com) that offered draft-writing assistance.

FUNDING INFORMATION

The research is supported by the fund of The Grants of Peak Climbing Project of Foshan Hospital of Traditional Chinese Medicine (CN) (No. 202000190, No. 202200087, No. 2022000189, No. 202000206).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data supporting the findings of this study can be found in the article and Supporting Information.

ORCID

Danqing Guo https://orcid.org/0000-0002-9538-2237

REFERENCES

1. Lin CC, Li Q, Williams CM, et al. The economic burden of guideline-recommended first line care for acute low back pain. *Eur Spine J*. 2018;*27*(1):109-116.
2. Jensen CE, Riis A, Petersen KD, et al. Economic evaluation of an implementation strategy for the management of low back pain in general practice. *Pain*. 2017;*158*(5):891-899.
3. Luoma K, Riihimaki H, Luukkonen R, et al. Low back pain in relation to lumbar disc degeneration. *Spine*. 2000;*25*(4):487-492.
4. Deyo RA, Cherkin D, Conrad D, et al. Cost, controversy, crisis: low back pain and the health of the public. *Annu Rev Public Health*. 1991;*12*:141-156.
5. Li Y, He XN, Li C, et al. Identification of candidate genes and microrna s for acute myocardial infarction by weighted gene coexpression network analysis. *BioMed Res Int*. 2019;*2019*:5742608.
6. Li H, Li W, Zhang L, et al. Comprehensive network analysis identified Sirt7, Ntrk2, and Chi3L1 as new potential markers for intervertebral disc degeneration. *J Oncol*. 2022;*2022*:4407541.
7. Wang N, Liu X, Fang X, et al. Identification of Smim1 and Sez6L2 as potential biomarkers for genes associated with intervertebral disc degeneration in pyroptosis. *Dis Markers*. 2022;*2022*:9515571.
8. Ji SC, Han N, Liu Y, et al. Identification of genes associated with disc degeneration using bioinformatics. *Biotech Histochem*. 2015;*90*(5):353-360.
9. Langfelder P, Horvath S. Wgcna: an R package for weighted correlation network analysis. *BMC Bioinform*. 2008;*9*:559.
10. Ma X, Su J, Wang B, et al. Identification of characteristic genes in whole blood of intervertebral disc degeneration patients by weighted gene coexpression network analysis (Wgcna). *Comput Math Methods Med*. 2022;*2022*:6609901.
11. Zhang Y, Han S, Kong M, et al. Single-cell RNA-seq analysis identifies unique chondrocyte subsets and reveals involvement of ferroptosis in human intervertebral disc degeneration. *Osteoarthritis Cartilage*. 2021.
12. Fernandes LM, Khan NM, Trochez CM, et al. Single-cell RNA-seq identifies unique transcriptional landscapes of human nucleus pulposus and annulus fibrosus cells. *Sci Rep*. 2020;*10*(1):15263.
13. Ye F, Lyu FJ, Wang H, et al. The involvement of immune system in intervertebral disc herniation and degeneration. *JOR Spine*. 2022;*5*(1):e1196.
14. Koroth J, Buko EO, Abbott R, et al. Macrophages and intervertebral disc degeneration. *Int J Mol Sci*. 2023;*24*(2).
15. Bermudez-Lekerika P, Crump KB, Tseranidou S, et al. Immuno-modulatory effects of intervertebral disc cells. *Front Cell Dev Biol*. 2022;*10*:924692.
16. Phillips KL, Chiverton N, Michael AL, et al. The cytokine and chemokine expression profile of nucleus pulposus cells: implications for

degeneration and regeneration of the intervertebral disc. *Arthritis Res Ther*. 2013;*15*(6):R213.
17. Risbud MV, Shapiro IM. Role of cytokines in intervertebral disc degeneration: pain and disc content. *Nat Rev Rheumatol*. 2014;*10*(1): 44-56.
18. Gruber HE, Hoelscher GL, Ingram JA, et al. Genome-wide analysis of pain-, nerve- and neurotrophin-related gene expression in the degenerating human annulus. *Mol Pain*. 2012;*8*:63.
19. Gruber HE, Hoelscher G, Loeffler B, et al. Prostaglandin E1 and misoprostol increase epidermal growth factor production in 3D-cultured human annulus cells. *Spine J*. 2009;*9*(9):760-766.
20. Thompson JP, Pearce RH, Schechter MT, et al. Preliminary evaluation of a scheme for grading the gross morphology of the human intervertebral disc. *Spine*. 1990;*15*(5):411-415.
21. Griffith JF, Wang YX, Antonio GE, et al. Modified Pfirrmann grading system for lumbar intervertebral disc degeneration. *Spine*. 2007;*32* (24):E708-E712.
22. Smyth GK. Limma: linear models for microarray data. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer; 2005:397-420.
23. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis [J]. *Stat Appl Genet Mol Biol*. 2005;*4*: Article17.
24. Barabasi AL, Bonabeau E. Scale-free networks. *Sci Am*. 2003;*288*(5): 60-69.
25. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. *Bioinformatics*. 2008;*24*(5):719-720.
26. Tang J, Kong D, Cui Q, et al. Prognostic genes of breast cancer identified by gene co-expression network analysis. *Front Oncol*. 2018; *8*:374.
27. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;*16*(5): 284-287.
28. Walter W, Sanchez-Cabo F, Ricote M. Goplot: an R package for visually combining expression data with functional analysis. *Bioinformatics*. 2015;*31*(17):2912-2914.
29. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J R Stat Soc Series B*. 1996;*58*:267-288.
30. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010;*33*(1): 1-22.
31. Robin X, Turck N, Hainard A, et al. proc: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform*. 2011; *12*:77.
32. Kepler CK, Ponnappan RK, Tannoury CA, et al. The molecular basis of intervertebral disc degeneration. *Spine J*. 2013;*13*(3):318-330.
33. Le Maitre CL, Pockert A, Buttle DJ, et al. Matrix synthesis and degradation in human intervertebral disc degeneration. *Biochem Soc Trans*. 2007;*35*(Pt 4):652-655.
34. Liao Z, Luo R, Li G, et al. Exosomes from mesenchymal stem cells modulate endoplasmic reticulum stress to protect against nucleus pulposus cell death and ameliorate intervertebral disc degeneration in vivo. *Theranostics*. 2019;*9*(14):4084-4100.
35. Luo R, Li G, Zhang W, et al. O-GlcNAc transferase regulates intervertebral disc degeneration by targeting FAM134B-mediated ER-phagy. *Exp Mol Med*. 2022;*54*(9):1472-1485.
36. So JS. Roles of endoplasmic reticulum stress in immune responses. *Mol Cells*. 2018;*41*(8):705-716.
37. Li A, Song NJ, Riesenberg BP, et al. The emerging roles of endoplasmic reticulum stress in balancing immunity and tolerance in health and diseases: Mechanisms and opportunities. *Front Immunol*. 2019; *10*:3154.
38. Li W, Zhao Y, Wang Y, et al. Deciphering the sequential changes of monocytes/macrophages in the progression of IDD with longitudinal approach using single-cell transcriptome. *Front Immunol*. 2023;*14*: 1090637.
39. Kratofil RM, Kubes P, Deniset JF. Monocyte conversion during inflammation and injury. *Arterioscler Thromb Vasc Biol*. 2017;*37*(1):35-42.
40. Phillips KL, Cullen K, Chiverton N, et al. Potential roles of cytokines and chemokines in human intervertebral disc degeneration: interleukin-1 is a master regulator of catabolic processes. *Osteoarthritis Cartilage*. 2015;*23*(7):1165-1177.
41. Nagasawa K, Higashi T, Hosokawa N, et al. Simultaneous induction of the four subunits of the Trap complex by ER stress accelerates ER degradation. *Embo Rep*. 2007;*8*(5):483-489.
42. Huang S, Zhong W, Shi Z, et al. Overexpression of signal sequence receptor γ predicts poor survival in patients with hepatocellular carcinoma. *Hum Pathol*. 2018;*81*:47-54.
43. Yan J, Wang ZH, Yan Y, et al. Rp11-156L14.1 regulates Ssr1 expression by competitively binding to miR-548ao-3p in hypopharyngeal squamous cell carcinoma. *Oncol Rep*. 2020;*44*(5):2080-2092.
44. Zhang W, Shen J, Wang Y, et al. Blood Ssr1: A possible biomarker for early prediction of Parkinson's Disease. *Front Mol Neurosci*. 2022;*15*: 762544.
45. Schweizer A, Ericsson M, Bachi T, et al. Characterization of a novel 63 kDa membrane protein. Implications for the organization of the ER-to-Golgi pathway. *J Cell Sci*. 1993;*104*(Pt 3):671-683.
46. Conrads TP, Tocci GM, Hood BL, et al. Ckap4/p63 is a receptor for the frizzled-8 protein-related antiproliferative factor from interstitial cystitis patients. *J Biol Chem*. 2006;*281*(49):37836-37843.
47. Razzaq TM, Bass R, Vines DJ, et al. Functional regulation of tissue plasminogen activator on the surface of vascular smooth muscle cells by the type-II transmembrane protein p63 (CKAP4). *J Biol Chem*. 2003;*278*(43):42679-42685.
48. Klopfenstein DR, Klumperman J, Lustig A, et al. Subdomain-specific localization of Climp-63 (p63) in the endoplasmic reticulum is mediated by its luminal alpha-helical segment. *J Cell Biol*. 2001;*153*(6): 1287-1300.
49. Pepin G, Perron MP, Provost P. Regulation of human Dicer by the resident Er membrane protein Climp-63. *Nucleic Acids Res*. 2012;*40* (22):11603-11617.
50. Li SX, Li J, Dong LW, et al. Cytoskeleton-associated protein 4, a promising biomarker for tumor diagnosis and therapy. *Front Mol Biosci*. 2020;*7*:552056.
51. Kimura H, Fumoto K, Shojima K, et al. Ckap4 is a Dickkopf1 receptor and is involved in tumor progression. *J Clin Invest*. 2016;*126*(7):2689-2705.
52. Gladka MM, Molenaar B, De Ruiter H, et al. Single-cell sequencing of the healthy and diseased heart reveals cytoskeleton-associated protein 4 as a new modulator of fibroblasts activation. *Circulation*. 2018; *138*(2):166-180.
53. Shi Y, Jin X, Yang M, et al. CKAP4 contributes to the progression of vascular calcification (VC) in chronic kidney disease (CKD) by modulating Yap phosphorylation and MMP2 expression. *Cell Signal*. 2022; *93*:110270.
54. Ortega A, Roselló-Lletí E, TarazóN E, et al. Endoplasmic reticulum stress induces different molecular structural alterations in human dilated and ischemic cardiomyopathy. *PLoS One*. 2014;*9*(9):e107635.
55. Wang J, Thomas HR, Li Z, et al. Puma, noxa, p53, and p63 differentially mediate stress pathway induced apoptosis. *Cell Death Dis*. 2021; *12*(7):659.
56. Pyati UJ, Gjini E, Carbonneau S, et al. p63 mediates an apoptotic response to pharmacological and disease-related ER stress in the developing epidermis. *Dev Cell*. 2011;*21*(3):492-505.
57. Ji Q, Zheng Y, Zhang G, et al. Single-cell RNA-seq analysis reveals the progression of human osteoarthritis. *Ann Rheum Dis*. 2019;*78*(1): 100-110.
58. Miosge N, Waletzko K, Bode C, et al. Light and electron microscopic in-situ hybridization of collagen type I and type II mRNA in the

fibrocartilaginous tissue of late-stage osteoarthritis. *Osteoar Cartil*. 1998;*6*(4):278-285.
59. Zaucke F, Dinser R, Maurer P, et al. Cartilage oligomeric matrix protein (Comp) and collagen Ix are sensitive markers for the differentiation state of articular primary chondrocytes. *Biochem J*. 2001;*358*(Pt 1):17-24.
60. Chan DD, Li J, Luo W, et al. Pirfenidone reduces subchondral bone loss and fibrosis after murine knee cartilage injury. *J Orthop Res*. 2018;*36*(1):365-376.
61. Matthews JL, Chung M, Matyas JR. Indirect injury stimulates scar formation-adaptation or pathology? *Connect Tissue Res*. 2004;*45*(2):94-100.
62. Seol D, Yu Y, Choe H, et al. Effect of short-term enzymatic treatment on cell migration and cartilage regeneration: in vitro organ culture of bovine articular cartilage. *Tissue Eng Part A*. 2014;*20*(13-14):1807-1814.
63. Seol D, Mccabe DJ, Choe H, et al. Chondrogenic progenitor cells respond to cartilage injury. *Arthritis Rheum*. 2012;*64*(11):3626-3637.
64. Joos H, Wildner A, Hogrefe C, et al. Interleukin-1 beta and tumor necrosis factor alpha inhibit migration activity of chondrogenic progenitor cells from non-fibrillated osteoarthritic cartilage. *Arthritis Res Ther*. 2013;*15*(5):R119.
65. Koelling S, Kruegel J, Irmer M, et al. Migratory chondrogenic progenitor cells from repair tissue during the later stages of human osteoarthritis. *Cell Stem Cell*. 2009;*4*(4):324-335.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Guo, D., Zeng, M., Yu, M., Shang, J., Lin, J., Liu, L., Yang, K., & Cao, Z. (2024). SSR1 and CKAP4 as potential biomarkers for intervertebral disc degeneration based on integrated bioinformatics analysis. *JOR Spine*, *7*(1), e1309. https://doi.org/10.1002/jsp2.1309